LEARNING MADE EASY

2nd Edition

Predictive Analytics

for dummies

A Wiley Brand

Real-world tips for creating business value

Details on modeling, data clustering, and more

Enterprise use cases to help you get started

Anasse Bari, Ph.D.
Mohamed Chaouchi
Tommy Jung

Predictive Analytics

for dummies®

A Wiley Brand

2nd edition

by Anasse Bari, Ph.D.,
Mohamed Chaouchi, and
Tommy Jung

for dummies®
A Wiley Brand

Predictive Analytics For Dummies®, 2nd Edition

Published by: **John Wiley & Sons, Inc.,** 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Published simultaneously in Canada

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: THE PUBLISHER AND THE AUTHOR MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION WARRANTIES OF FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES OR PROMOTIONAL MATERIALS. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR EVERY SITUATION. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING LEGAL, ACCOUNTING, OR OTHER PROFESSIONAL SERVICES. IF PROFESSIONAL ASSISTANCE IS REQUIRED, THE SERVICES OF A COMPETENT PROFESSIONAL PERSON SHOULD BE SOUGHT. NEITHER THE PUBLISHER NOR THE AUTHOR SHALL BE LIABLE FOR DAMAGES ARISING HEREFROM. THE FACT THAT AN ORGANIZATION OR WEBSITE IS REFERRED TO IN THIS WORK AS A CITATION AND/OR A POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE AUTHOR OR THE PUBLISHER ENDORSES THE INFORMATION THE ORGANIZATION OR WEBSITE MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. FURTHER, READERS SHOULD BE AWARE THAT INTERNET WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit https://hub.wiley.com/community/support/dummies.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at http://booksupport.wiley.com. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2016951998

ISBN 978-1-119-26700-3 (pbk); 978-1-119-26701-0 (epub); 978-1-119-26702-7 (epdf)

Manufactured in the United States of America

10 9 8 7 6 5 4 3 2 1

Table of Contents

Introduction

Predictive Analytics is the art and science of using data to make better informed decisions. Predictive analytics helps you uncover hidden patterns and relationships in your data that can help you predict with greater confidence what may happen in the future, and provide you with valuable, actionable insights for your organization.

About This Book

Our goal was to make this complex subject as practical as possible, in a way that appeals to everyone from technical experts to non-technical level business strategists.

The subject is complex because it is not really just one subject. It is the combination of at least a few multifaceted fields: data mining, statistics, and mathematics.

Data mining requires an understanding of machine learning and information retrieval. On top of this, mathematics and statistics must be applied to your business domain; be it marketing, actuary service, fraud, crime, or banking.

Most of the current materials on predictive analytics are pretty difficult to read if you don't already have a background in some of the aforementioned subjects. They are filled with complex mathematical equations and modeling techniques. Or, they are at a high level with specific use cases but with little guidance regarding implementation. We include both, while trying to keep a wide spectrum of readers engaged.

The focus of this book is developing a roadmap for implementing predictive analytics within your organization. Its intended audience is the larger community of business managers, business analysts, data scientists, and information technology professionals.

Maybe you are a business manager and you have heard the buzz about predictive analytics. Maybe you've been working with data mining and you want to add predictive analytics to your skill set. Maybe you know R or Python, but you're totally new to predictive analytics. If this sounds like you, then this book will be a good fit. Even if you have no experience analyzing data, but want or need to derive greater value from your organization's data, you can also find something of value in this book.

Foolish Assumptions

Without oversimplifying, we have tried to explain technical concepts in nontechnical terms, tackling each topic from the ground up.

Even if you are an experienced practitioner, you should find something new, and at the very least, you will gain validation for what you already know, and guidance for establishing best practices.

We also hope to have contributed a few concepts and ideas for the very first time in a major publication like this. For example we explain how you can apply biologically inspired algorithms to predictive analytics.

We assume that the reader will not be a programmer. The code presented in this book is very brief and easy to follow. Readers of all programming levels will benefit from this book, because it is more about learning the process of predictive analytics rather than learning a programming language.

Icons Used in This Book

The following icons in the margins indicate highlighted material that we think could be of interest to you. Next, we describe the meaning of each icon that is used in this book.

TIP

The tips are ideas we would like you to take note of. This is usually practical advice you can apply for that given topic.

WARNING

This icon is rarely used in this book. We may have used it only once or twice in the entire book. The intent is to save you time by bringing to your attention some common pitfalls that you are better off avoiding.

TECHNICAL STUFF

We have made sincere efforts to steer away from the technical stuff. But when we have no choice we make sure to let you know. So if you don't care too much about the technical stuff you can easily skip this part and you won't miss much. If the technical stuff is your thing, then you may find these sections fascinating.

REMEMBER

This is something we would like you to take a special note of. This is a concept or idea we think is important for you know and remember. An example of this would be a best practice we think it is noteworthy.

Beyond the Book

A lot of extra content that is not in this book is available at `www.dummies.com`. Go online to find the following:

- **The Cheat Sheet for this book is at**

  `www.dummies.com/cheatsheet/predictiveanalytics`

  Here you'll find the necessary steps needed to build a predictive analytics model and some cases studies of predictive analytics.

- **Updates to this book, if we have any, are also available at**

  `www.dummies.com/extras/predictiveanalytics`

Where to Go from Here

Let's start making some predictions! You can apply predictive analytics to virtually every business domain. Right now there is explosive growth in predictive analytics' market, and this is just the beginning. The arena is wide open, and the possibilities are endless.

1

Getting Started with Predictive Analytics

IN THIS PART . . .

Exploring predictive analytics

Identifying uses

Classifying data

Presenting information

IN THIS CHAPTER

Explaining the building blocks

Probing capabilities

Surveying the market

Chapter 1
Entering the Arena

Predictive analytics is a bright light bulb powered by your data.

You can never have too much insight. The more you see, the better the decisions you make — and you never want to be in the dark. You want to see what lies ahead, preferably before others do. It's like playing the game "Let's Make a Deal" where you have to choose the door with the hidden prize. Which door do you choose? Door 1, Door 2, or Door 3? They all look the same, so it's just your best guess — your choice depends on you and your luck. But what if you had an edge — the ability to see through the keyhole? Predictive analytics can give you that edge.

Exploring Predictive Analytics

What would you do in a world where you know how likely you are to end up marrying your college roommate? Where you can predict what profession will best suit you? Where you can predict the best city and country for you to live in?

In short, imagine a world where you can maximize the potential of every moment of your life. Such a life would be productive, efficient, and powerful. You will (in effect) have superpowers — and a lot more spare time. Well, such a world may seem a little boring to people who like to take uncalculated risks, but not to a profit-generating organization. Organizations spend millions of dollars managing risk. And if there is something out there that helps them manage their risk, optimize their operations, and maximize their profits, you should definitely learn about it. That is the world of predictive analytics.

Mining data

Big data is the new reality. In fact, data is only getting bigger, faster, and richer. It's here to stay and you'd better capitalize on it.

Data is one of your organization's most valuable assets. It's full of hidden value, but you have to dig for it. *Data mining* is the discovery of hidden patterns of data through machine learning — and sophisticated algorithms are the mining tools. *Predictive analytics* is the process of refining that data resource, using business knowledge to extract hidden value from those newly discovered patterns.

Data mining + business knowledge = predictive analytics => value

Today's leading organizations are looking at their data, examining it, and processing it to search for ways to better understand their customer base, improve their operations, outperform their competitors, and better position themselves in the marketplace. They are looking into how they can use that information to increase their market share and sharpen their competitive edge. How can they drive better sales and more effectively target marketing campaigns? How can they better serve their customers and meet their needs? What can they do to improve the bottom line?

But these tools are useful in realms beyond business. As one major example, government law enforcement agencies are asking questions related to crime detection and prevention. Is this a person of interest? Will this criminal be a repeat offender? Where will the next crime happen?

Other industries, notably those with financial responsibility, could use a trustworthy glimpse into the future. Companies are trying to know whether the transaction they're currently processing is fraudulent, whether an insurance claim is legitimate, whether a credit card purchase is valid, whether a credit applicant is worthy of credit . . . the list goes on.

Governments, companies, and individuals are (variously) looking to spot trends in social movements, detect emerging healthcare issues and disease outbreaks, uncover new fashion trends, or find that perfect lifetime partner.

These — and plenty more — business and research questions are topics you can investigate further to find answers to by mining the available data and building predictive analytics models to guide future decisions.

Data + predictive analytics = light.

Highlighting the model

A *model* is a mathematical representation of an object or a process. We build models to simulate real-world phenomena as a further investigative step, in hopes of understanding more clearly what's really going on. For example, to model our customers' behavior, we seek to mimic how our customers have been navigating through our websites:

- Which products did they look at before they made a purchase?
- Which pages did they view before making that purchase?
- Did they look at the products' descriptions?
- Did they read users' reviews?
- How many reviews did they read?
- Did they read both positive and negative reviews?
- Did they purchase something else in addition to the product they came looking for?

We collect all that data from past occurrences. We look at those historical transactions between our company and our customers — and try to make consistent sense of them. We examine that data and see whether it holds answers to our questions. Collecting that data — with particular attention to the breadth and depth of the data, its quality level, and its predictive value — helps to form the boundaries that will define our model and its outputs.

This process isn't to be confused with just reporting on the data; it's also different from just visualizing that data. Although those steps are vital, they're just the beginning of exploring the data and gaining a usable understanding of it.

We go a lot deeper when we're talking about developing predictive analytics. In the first place, we need to take a threefold approach:

- Thoroughly understand the business problem we're trying to solve.
- Obtain and prepare the data we want our model to work with.
- Run statistical analysis, data-mining, and machine-learning algorithms on the data.

In the process, we have to look at various *attributes* — data points we think are relevant to our analysis. We'll run several *algorithms*, which are sets of mathematical instructions that get machines to do problem-solving. We keep running through possible combinations of data and investigate what-if scenarios. Eventually we build our model, find our answers, and prepare to deploy that model and reap its benefits.

What does a model look like? Well, in programming terms, a predictive analytics model can be as simple as a few `if ... then` statements that tell the machine, "If this condition exists, then perform this action."

Here are some simple rule-based trading models:

- If it's past 10:00a.m. ET and the market is up, then buy 100 shares of XYZ stock.
- If my stock is up by 10 percent, then take profits.
- If my portfolio is down by 10 percent, then exit my positions.

Here's a simple rule-based recommender system (for more about recommender systems, see Chapter 2):

- If a person buys a book by this author, then recommend other books by the same author.
- If a person buys a book on this topic, then recommend other books on the same and related topic.
- If a person buys a book on this topic, then recommend books that other customers have purchased when they bought this book.

Adding Business Value

In an increasingly competitive environment, organizations always need ways to become more competitive. Predictive analytics found its way into organizations as one such tool. Using technology in the form of machine-learning algorithms, statistics, and data-mining techniques (Chapter 3 shows how these disciplines differ and overlap), organizations can uncover hidden patterns and trends in their data that can aid in operations and strategy and help fulfill critical business needs.

Embedding predictive analytics in operational decisions improves return on investment because organizations spend less time dealing with low-impact, low-risk operational decisions. Employees can focus more of their time on high-impact, high-risk decisions. For example, most standard insurance claims can be automatically paid out. However, when the predictive model comes across a claim that's unusual (an outlier), or when the claim exhibits the same pattern as a fraudulent claim, the system can flag the claim automatically and send it to the appropriate person to take action.

By using predictive analytics to predict a future event or trend, the company can create a strategy to position itself to take advantage of that insight. If your predictive model is telling you (for example) that the trend in fashion is toward black turtlenecks, you can take appropriate actions to design more black-colored turtlenecks or design more accessories to go with the fashionable item. Tom Khabaza's Eighth Law of Data Mining summarizes this perfectly. "The value of a predictive model arises in two ways:

> The model's predictions drive improved (more effective) action, and
>
> The model delivers insight (new knowledge) which leads to improved strategy."

Endless opportunities

Organizations around the world are striving to improve, compete, and be lean. They're looking to make their planning process more agile. They're investigating how to manage inventories and optimize the allocations of their human resources to best advantage. They're looking to act on opportunities as they arise in real time.

Predictive analytics can make all those goals more reachable. The domains to which predictive analytics can be applied are unlimited; the arena is wide open and everything is fair game. Let the mining start. Let the analysis begin.

Go to your analytics team and have them mine the data you've accumulated or acquired, with an eye toward finding an advantageous niche market for your product; innovate with data. Ask the team to help you gain confidence in your decision-making and risk management.

Albert Einstein once said, "Know where to find information and how to use it; that is the secret of success." If that's the secret to success, then you will succeed by using predictive analytics: The information is in your data and data mining will find it. The rest of the equation relies on your business knowledge of how to interpret that information — and ultimately use it to create success.

Finding value in data equals success. Therefore we can rewrite our predictive analytics equation as

Data mining + business knowledge = predictive analytics => success

Empowering your organization

Predictive analytics empowers your organization by providing three advantages:

- Vision
- Decision
- Precision

Vision

Predictive analytics will lead you to see what is invisible to others — in particular, useful patterns in your data.

Predictive analytics can provide you with powerful hints to lend direction to the decisions you're about to make in your company's quest to retain customers, attract more customers, and maximize profits. Predictive analytics can go through a lot of past customer data, associate it with other pieces of data, and assemble all the pieces in the right order to solve that puzzle in various ways, including

- Categorizing your customers and speculating about their needs.
- Knowing your customers' wish lists.
- Guessing your customers' next actions.
- Categorizing your customers as loyal, seasonal, or wandering.

Knowing this type of information beforehand shapes your strategic planning and helps optimize resource allocation, increase customer satisfaction, and maximize your profits.

Decision

A well-made predictive analytics model provides analytical results free of emotion and bias. The model uses mathematical functions to derive forward insights from numbers and text that describe past facts and current information. The model provides you with consistent and unbiased insights to support your decisions.

Consider the scenario of a typical application for a credit card: The process takes a few minutes; the bank or agency makes a quick, fact-based decision on whether to extend credit, and is confident in their decision. The speed of that transaction is possible thanks to predictive analytics, which predicted the applicant's creditworthiness.

Precision

Imagine having to read a lot of reports, derive insights from the past facts buried in them, go through rows of Excel spreadsheets to compare results, or extract information from a large array of numbers. You'd need a staff to do these time-consuming tasks. With predictive analytics, you can use automated tools to do the job for you — saving time and resources, reducing human error, and improving precision.

For example, you can focus targeted marketing campaigns by examining the data you have about your customers, their demographics, and their purchases. When you know precisely which customers you should market to, you can zero in on those most likely to buy.

Starting a Predictive Analytic Project

For the moment, let's forget about algorithms and higher math; predictions are used in every aspect of our lives. Consider how many times you have said (or heard people say), "I told you that was going to happen."

When you want to predict a future event with any accuracy, however, you'll need to know the past and understand the current situation. Doing so entails several processes:

- Extract the facts that are currently happening.
- Distinguish present facts from those that just happened.
- Derive possible scenarios that could happen.
- Rank the scenarios according to how likely they are to happen.

Predictive analytics can help you with each of these processes, so that you know as much as you can about what has happened and can make better-informed decisions about the future.

Companies typically create predictive analytics solutions by combining three ingredients:

- Business knowledge
- Data-science team and technology
- The data

Though the proportion of the three ingredients will vary from one business to the next, all are required for a successful predictive analytic solution that yields actionable insights.

Business knowledge

Because any predictive analytics project is started to fulfill a business need, business-specific knowledge and a clear business objective are critical to its success. Ideas for a project can come from anyone within the organization, but it's up to the leadership team to set the business goals and get buy-in from the needed departments across the whole organization.

Be sure the decision-makers in your team are prepared to act. When you present a prototype of your project, it needs an in-house champion — someone who's going to push for its adoption.

The leadership team or domain experts must also set clear *metrics* — ways to quantify and measure the outcome of the project. Appropriate metrics keep the departments involved clear about what they need to do, how much they need to do, and whether what they're doing is helping the company achieve its business goals.

The *business stakeholders* are those who are most familiar with the domain of the business. They'll have ideas about which correlations — relationships between features — of data work and which don't, which variables are important to the model, and whether you should create new variables — as in derived features or attributes — to improve the model.

Business analysts and other domain experts can analyze and interpret the patterns discovered by the machines, making useful meaning out of the data patterns and deriving actionable insights.

This is an *iterative* (building a model and interpreting its findings) process between business and science. In the course of building a predictive model, you have to try successive versions of the model to improve how it works (which is what data experts mean when they say *iterate the model over its lifecycle*). You might go through a lot of revisions and repetitions before you can prove that your model is bringing real value to the business. Even after the predictive models are deployed, the business must monitor the results, validate the accuracy of the models and improve upon the models as more data is being collected.

Data-science team and technology

The technology used in predictive analytics will include at least some (if not all) of these capabilities:

- Data mining
- Statistics
- Machine-learning algorithms
- Software tools to build the model

The business people needn't understand the details of all the technology used or the math involved — but they should have a good handle on the process that model represents, and on how it integrates with the overall infrastructure of your organization. Remember, this is a collaborative process; the data scientists and business people must work closely together to build the model.

By the same token, providing a good general grasp of business knowledge to the data scientists gives them a better chance at creating an accurate predictive model, and helps them deploy the model much more quickly. After the model is deployed, the business can start evaluating the results right away — and the teams can start working on improving the model. Through testing, the teams will learn together what works and what doesn't.

The combination of business knowledge, data exploration, and technology leads to a successful deployment of the predictive model. So the overall approach is to develop the model through successive versions and make sure the team members have enough knowledge of both the business and the data science that everyone is on the same page.

Some analytical tools — specialized software products — are advanced enough that they require people with scientific backgrounds to use them; others are simple enough that any business person within the organization can use them. Selecting the right tool(s) is also a decision that must be taken very carefully. Every company will have different needs and not any one tool can address all those needs. But one thing is certain; every company will have to use some sort of tool to do predictive analytics.

Selecting the right software product for the job depends on such factors as

- The cost of the product
- The complexity of the business problem

- The complexity of data
- The source(s) of the data
- The velocity of the data (the speed by which the data changes)
- The people within the organization who will use the product

The Data

All else being equal, you'd expect a person who has more experience to be better at doing a job, playing a game, or whatever than someone who has less experience. That same thinking can be applied to an organization. If you imagine an organization as a person, you can view the organization's data as its equivalent of experience. By using that experience, you can make more insightful business decisions and operate with greater efficiency. Such is the process of turning data into business value with predictive analytics.

It's increasingly clear that data is a vital asset for driving the decision-making process quick, realistic answers and insights. Predictive analytics empower business decisions by uncovering opportunities such as emerging trends, markets, or customers before the competition.

Data can also present a few challenges in its raw form. It can be distributed across multiple sources, mix your own data with third-party data, and otherwise make the quality of incoming data too messy to use right away. Thus you should expect your data scientists to spend considerable time exploring your data and preparing it for analysis. This process of *data cleansing* and *data preparation* involves spotting missing values, duplicate records, and outliers, generating derived values, and normalization. (For more about these processes, see Chapters 9 and 15.)

Big data has its own challenging properties that include volume, velocity, and variety: In effect, too much of it comes in too fast, from too many places, in too many different forms. Then the main problem becomes separating the relevant data from the noise surrounding it.

In such a case, your team has to evaluate the state of the data and its type, and choose the most suitable algorithm to run on that data. Such decisions are part of an exploration phase in which the data scientists gain intimate knowledge of your data while they're selecting which attributes have the most predictive power.

Ongoing Predictive Analytics

Predictive analytics should never be about implementing one project or two, even if those two projects are very successful. It should be an ongoing process that feeds into, and is enforced by, the governing body overseeing strategy and operational planning at your organization.

You should put data at the forefront of the decision-making process at your organization. Data must support any major initiatives. After collecting and acquiring all relevant data, have your data-science team make sense of it, and propose a way forward based on their findings. The outcomes of these efforts should reach the entire organization by fostering a cultural change that embraces the analytical work as an accepted way to make informed decisions.

Your work on predictive models doesn't stop at the moment you deploy them. That only gets your foot in the door. You should actually be constantly looking for ways to improve that model. Models tend to decay over time. So refreshing the model is a necessary step in building predictive analytics solutions. The model should be undergoing continuous improvement.

Additionally, you may have several models deployed, and each one of them may have undergone several revisions. In such case, it's imperative to have processes in place to manage the models' lifecycle, overseeing the creation, updating, and retiring of each model. Depending on the line of the business you're in, you may need to audit all changes and be very granular in your documentation of all steps involved in this process.

Your belief in the promise of predictive analytics should never stop you from questioning the results of a predictive analytics project. You can't just go ahead and implement blindly. You should make sure that the results make sense businesswise. Also, when the results are too good to be true, they probably are. Verify the correctness and accuracy of all steps followed to generate those models. Scrutinizing the models' results and asking the hard questions will only further your confidence in the decisions you will finally make based on those findings.

Sometimes the results of a predictive analytics project can be so obvious that business stakeholders may dismiss them altogether on the pretense that "we already knew that". Keep in mind, however, that making the effort to thoroughly understand the outputs of a model can be rewarding, no matter how obvious the results may seem at first.

When (for example) a model shows you that 90 percent of your customers are urban, and are between the ages of 25 and 45, the results may seem obvious. You may feel you wasted time and resources to only find out what you already knew.

It may be far more important, however, to ask *what the other 10% are made of.* How can you increase *their* percentages? You may need to build a new model to find out more about that segment of your customers. Or you may want to learn more about what attracts 90 percent of your customers to your product.

Building predictive analytics models should be an ongoing process and the results should be shared across the organization. You should always be looking to improve your models; never shy away from both experimenting and asking the hard questions. With relevant data, a talented data-science team, and the buy-in from the business stakeholders, the possibilities are endless.

Forming Your Predictive Analytics Team

A successful predictive analytics team blends the necessary skills and attitude. We'll bet you can find them in your organization.

Hiring experienced practitioners

The data-science team should be composed of experienced practitioners. Experienced data scientists know their way around data. They know what models work best for which business problems and data types.

It should be required for your data science team to have members with professional knowledge and proven experience in statistics, data mining, and machine learning. These three disciplines should be mandatory for any data science team; the idea is that these skills must exist within the team, not necessarily that every team member needs all three. However, hiring team members from diverse backgrounds can spice up and enrich your team. Other experiences and knowledge of other disciplines can make the overall team more rounded and can broaden its horizons.

Among the team members you hire to join your data-science team should be data scientists who have knowledge of your specific business domain. That business knowledge could come from past experience working on projects in your business domain or in fields or related to it. The more the team members know about your line of business, the easier it will be for them to work with your data and build analytical solutions.

There are many powerful tools provided by many vendors, in addition to great open source tools available to you. Your team members should have working knowledge with these tools. This will facilitate the life cycle of building analytical solutions. Also, that knowledge will facilitate collaboration across the team members and between business analysts and data scientists.

Demonstrating commitment and curiosity

Senior management should show their commitment to the analytical efforts. They should meet with the team members and follow the progress of their projects. They should allocate time to be briefed about the projects, their progress, and their final findings.

Your data science team members should believe in the mission and be committed to finding answers to the business questions they are after. Keeping the team members motivated and engaged will help allow them to thrive to deliver the best solutions. Team members should be curious and excited to achieve the business goals.

Your team members should be able to communicate their findings in a language understood by your business stakeholders. When the team members are able to communicate with the business stakeholders, they will be able to gather support for the new solutions and get the necessary buy-in. This is especially important when the business users will need to change the way they have been doing their work when they start applying the new findings.

The team members should be curious, always asking questions, and trying to learn as much as they can about their projects. By not shying away from asking the toughest questions about the data, methods used and models outputs, and not shying away from trying even the wackiest scenarios, the team members will deliver optimal solutions.

Collaboration among team members and across the rest of the organization is important to the success of these projects. Team members should be able to help each other and answer each other's questions. Also they should be able to share the results and get immediate feedback.

Surveying the Marketplace

Big data and predictive analytics are bringing equally big changes to academia, the job market, and virtually every competitive company out there. Everybody will feel the impact. The survivors will treat it as an opportunity.

Responding to big data

Numerous universities offer certificates and master's degrees in predictive analytics or big-data analytics; some of these degree programs have emerged within the past year or two. This reflects the amazing growth and popularity of this field. The occupation of "data scientist" is now being labeled as one of the sexiest jobs in America by popular job journals and websites.

This demand in job growth is expected to grow; the projection is that job positions will outnumber qualified applicants. Some universities are shifting their program offerings to take advantage of this growth and attract more students. Some offer analytics programs in their business schools; while others provide similar offerings in their science and engineering schools. Like the real-world applications that handle big data and predictive analytics, the discipline that makes use of them spans departments — you can find relevant course offerings in business, mathematics, statistics, and computer science. The result is the same: more attractive and relevant degree programs for today's economy, and more students looking for a growing occupational field.

Working with big data

We read stories every day about how a hot new company is springing up using predictive analytics to solve specific problems — from predicting what you will do at every turn throughout the day to scoring how suitable you are as a boyfriend. Pretty wild. No matter how outrageous the concept, someone seems to be doing it. People and companies do it for a straightforward reason: There is a market for it. There is a huge demand for social analytics, people analytics, *everything data* analytics.

Statisticians and mathematicians — whose primary task once consisted primarily of sitting at desks and crunching numbers for drug and finance companies — are now in the forefront of a data revolution that promises to predict nearly everything about nearly everyone — including you.

So why are we witnessing this sudden shift in analytics? After all, mathematics, statistics and their derivatives, computer science, machine learning, and data mining have been here for decades. In fact, most of the algorithms in use today to develop predictive models were created decades ago. The answer has to be "data" — lots of it.

We gather and generate huge amounts of data every day. Only recently have we been able to mine this data effectively. Processing power and data storage have increased exponentially while getting faster and cheaper. We've figured out how to use computer hardware to store and process large amounts of data.